

Klos, Silke; Henke, Christian; Kieren, Corinna; Walpuski, Maik; Sumfleth, Elke
Naturwissenschaftliches Experimentieren und chemisches Fachwissen - zwei verschiedene Kompetenzen

Zeitschrift für Pädagogik 54 (2008) 3, S. 304-321



Quellenangabe/ Reference:

Klos, Silke; Henke, Christian; Kieren, Corinna; Walpuski, Maik; Sumfleth, Elke: Naturwissenschaftliches Experimentieren und chemisches Fachwissen - zwei verschiedene Kompetenzen - In: Zeitschrift für Pädagogik 54 (2008) 3, S. 304-321 - URN: urn:nbn:de:0111-opus-43535 - DOI: 10.25656/01:4353

<https://nbn-resolving.org/urn:nbn:de:0111-opus-43535>

<https://doi.org/10.25656/01:4353>

in Kooperation mit / in cooperation with:

BELTZ

<http://www.beltz.de>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.
Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.
This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipt.de
Internet: www.pedocs.de

Inhaltsverzeichnis

Thementeil: Lehr- und Lernprozesse im naturwissenschaftlichen Unterricht

Hans E. Fischer

Lehr- und Lernprozesse im naturwissenschaftlichen Unterricht.

Einführung in den Thementeil 301

Silke Klos/Christian Henke/Corinna Kieren/Maik Walpuski/Elke Sumfleth

Naturwissenschaftliches Experimentieren und chemisches Fachwissen –

zwei verschiedene Kompetenzen 304

Georg Trendel/Rainer Wackermann/Hans E. Fischer

Lernprozessorientierte Fortbildung von Physiklehrern 322

Isabell van Ackeren/Rainer Block/Klaus Klemm/Harry Kullmann/Frank Sprütten

Schulkultur als Kontext naturwissenschaftlichen Lernens – Allgemeine und

fachspezifische explorative Analysen 341

Joachim Wirth/Hubertina Thillmann/Josef Künsting/

Hans E. Fischer/Detlev Leutner

Das Schülerexperiment im naturwissenschaftlichen Unterricht – Bedingungen

der Lernförderlichkeit dieser Lehrmethode 361

Allgemeiner Teil

Frauke Stübiger/Peter H. Ludwig/Dorit Bosse

Problemorientierte Lehr-Lern-Arrangements in der Praxis –

Eine empirische Untersuchung zur Organisation und Gestaltung

fächerübergreifenden Unterrichts 376

Paul Walter/Achim Leschinsky

Überschätzte Helfer? Erwartungen an die Sozialpädagogik in der Schule 396

<i>Jörg Wittwer</i>	
Warum wirkt Nachhilfe? Hinweise aus der Forschung zum Einzelunterricht	416

Besprechungen

<i>Wolfgang Harder</i>	
Ulrich Herrmann (Hrsg.): In der Pädagogik etwas bewegen	433

Micha Brumlik

Johannes Bellmann: John Dewey naturalistische Pädagogik	
Fritz Bohnsack: John Dewey. Ein pädagogisches Portrait	
Martin Hartmann: Die Kreativität der Gewohnheit	
Klaus Prange (Hrsg.): Herbart und Dewey	
Douglas J. Simpson: John Dewey	
Robert Wentz: Demokratie am Scheideweg	435

Klaus Prange

Norbert Ricken (Hrsg.): Über die Verachtung der Pädagogik	438
---	-----

Jörg Zirfas

Andrea Sabisch: Inszenierung der Suche	441
--	-----

Dokumentation

Erziehungswissenschaftliche Habilitationen und Promotionen 2007	444
Pädagogische Neuerscheinungen	480

Beilagenhinweis:

Dieser Ausgabe der Z.f.Päd. liegen Prospekte des Juventa Verlag, Weinheim, und des Hogrefe Verlag, Göttingen, bei.

Silke Klos/Christian Henke/Corinna Kieren/Maik Walpuski/Elke Sumfleth

Naturwissenschaftliches Experimentieren und chemisches Fachwissen – zwei verschiedene Kompetenzen

***Zusammenfassung:** In diesem Beitrag wird über die Entwicklung und Evaluation eines Testverfahrens zur Messung Naturwissenschaftlich-experimenteller Arbeitsweisen, den so genannten NAW-Test, berichtet. Er wurde in zwei Varianten für die Jahrgangsstufe 7 und 12 entwickelt und in verschiedenen Studien eingesetzt, aus denen ausgewählte Ergebnisse berichtet werden. Die Ergebnisse bestätigen, dass mit dem NAW-Test ein valider Test vorliegt, der eine Kompetenz misst, welche sich von Fachwissen eindeutig unterscheidet.*

1. Einleitung

Die aktuelle deutsche bildungspolitische Situation fordert geeignete Aufgaben und Testinstrumente, mit deren Hilfe die Fähigkeiten in den durch die Bildungsstandards festgelegten Kompetenzbereichen erfasst werden können. Zur Überprüfung des Erreichens der Standards genügen Aufgaben für den mittleren Bildungsabschluss, für ein kontrolliertes Entwickeln der geforderten Leistungen müssen diese aber auch in anderen Jahrgangsstufen überprüft werden können. Zusätzlich werden Aufgaben benötigt, die nicht nur reines Fachwissen abprüfen, denn gerade in den Kompetenzbereichen, die *nicht* auf Fachwissen fokussieren, haben deutsche Schülerinnen und Schüler in den internationalen Vergleichsstudien unterdurchschnittliche Ergebnisse erzielt (Baumert u.a. 1997; Baumert u.a. 2001; PISA-Konsortium Deutschland: Prenzel u.a. 2004). In diesem Kontext ist an der Universität Duisburg-Essen ein Testverfahren entwickelt worden, mit dessen Hilfe experimentell-naturwissenschaftliche Arbeitsweisen erhoben werden können. Ausgehend von den ersten grundlegenden Arbeiten zur Testentwicklung für Schülerinnen und Schüler der Jahrgangsstufe 7 an Gymnasien wurden die Erkenntnisse auch für eine Testentwicklung für die Jahrgangsstufe 12 der gymnasialen Oberstufe genutzt. Da es sich hierbei um die Entwicklung eines völlig neuen Testinstruments handelt, muss dieses zunächst hinsichtlich klassischer Testgütekriterien evaluiert werden.

Im vorliegenden Beitrag wird, nach einem kurzen einführenden Überblick zu Bildungsstandards und Kompetenzen zunächst die theoretische Basis des entwickelten Testinstruments beschrieben, dann die Testgütekriterien unter besonderer Betonung der Validität und einiger Besonderheiten des Oberstufentests diskutiert und schließlich einige entscheidende Ergebnisse aus dem Einsatz dieser Testinstrumente dargestellt.

2. Bildungsstandards und Kompetenzen

Seit dem relativ schlechten Abschneiden deutscher Schülerinnen und Schüler in den internationalen Vergleichsstudien 2000 und 2003 (Baumert u.a. 2001; Baumert/Bos/Lehmann 2000a; Baumert/Bos/Lehmann 2000b; Prenzel u.a. 2004) wird in Deutschland verstärkt über das Bildungssystem und dessen Schwächen diskutiert. Der Deutsche Verein zur Förderung des mathematischen und naturwissenschaftlichen Unterrichts (2003, S. IV) fasst dies in der Aussage „In deutschen Schulen wird sehr viel gelehrt, einiges gelernt und zu wenig gekonnt“ zusammen. Als eine mögliche Ursache des schlechten Abschneidens deutscher Schülerinnen und Schüler wird die bislang externe Steuerung der schulischen und unterrichtlichen Praxis in Deutschland diskutiert. Helmke forderte dementsprechend bereits nach der TIMSS-Studie einen Wechsel von der Input- zur Outputsteuerung (Helmke 2000). Ein Vergleich mit anderen Nationen, wie zum Beispiel Finnland, zeigt, dass viele der Länder mit einem an Standards orientierten Bildungssystem bessere Ergebnisse innerhalb der Naturwissenschaften erzielen. Das bisher Versäumte wird in Deutschland nun nachgeholt. In den Richtlinien zur Entwicklung von allgemeinen Bildungsstandards für den mittleren Bildungsabschluss (Klieme u.a. 2003) wird zwischen Bildungszielen, Kompetenzanforderungen & -modellen und Aufgabenstellungen unterschieden. Die formulierten Bildungsziele müssen nach Kompetenzanforderungen und zu entwickelnden Kompetenzmodellen konkretisiert werden, wobei Kompetenzen vereinfacht als „die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können“ verstanden werden (Weinert 2001, S. 21). Folglich gilt es konkrete Aufgabenstellungen und Verfahren zur Erfassung der Schülerkompetenzen zu entwickeln.

Insgesamt legen „Bildungsstandards [...] fest, welche Kompetenzen die Kinder oder Jugendlichen bis zur einer bestimmten Jahrgangsstufe mindestens erworben haben sollen“ (Klieme u.a. 2003, S. 13). Während die Expertengruppe noch von Mindeststandards ausgeht und plant, auf diesen später Regelstandards aufzubauen, hat die KMK in den bisher vorliegenden Bildungsstandards für den mittleren Bildungsabschluss in den Fächern Mathematik, Biologie, Chemie, Physik und Englisch ausschließlich Regelstandards formuliert. Hierbei handelt es sich demnach um Kompetenzen, die in der Regel – und somit nicht von jedem Schüler oder jeder Schülerin – am Ende der 10. Jahrgangsstufe erreicht werden können. Als verbindliche Vorgabe für die Schulen finden sich in den Bildungsstandards konkret beschriebene, ausformulierte Kompetenzen, die überprüft werden sollen.

Für die jeweiligen Schulfächer werden verschiedene Kompetenzbereiche beschrieben. Die Bildungsstandards der naturwissenschaftlichen Fächer Biologie, Physik und Chemie unterscheiden jeweils vier Kompetenzbereiche *Fachwissen, Erkenntnisgewinnung, Kommunikation und Bewerten* (KMK 2004). Die gleichen Bereiche lassen sich ebenso in dem für das Unterrichtsfach Chemie entwickelten Kerncurriculum (Melle/

Parchmann/Sumfleth 2004) und in den Allgemeinen Prüfungsanforderungen für die Abiturprüfung für das Fach Chemie finden (KMK 2004). Die vier Kompetenzbereiche unterscheiden sich in den drei Naturwissenschaften nur durch den inhaltlichen Bezug zum jeweiligen Fach. Auf diese Weise werden fachspezifische und fächerübergreifende Kompetenzen berücksichtigt, die sowohl im Unterricht der drei Naturwissenschaften als auch in einem fächerübergreifenden integrierten naturwissenschaftlichen Unterricht erworben werden können. So können unterschiedliche Fachkontexte diskutiert werden und das Übertragen und Nutzen von Kompetenzen in verschiedenen Situationen gelernt werden. Eine Betrachtung der unterschiedlichen Kompetenzbereiche im Fach Chemie (KMK 2004) verdeutlicht, dass die vier Bereiche unterschiedlich stark miteinander verknüpft sind (Tab. 1).

Tab. 1: Kompetenzbereiche im Fach Chemie	
Fachwissen	Chemische Phänomene, Begriffe, Gesetzmäßigkeiten und Konzepte zuordnen
Erkenntnisgewinnung	Experimentelle und andere Untersuchungsmethoden sowie Modelle nutzen
Kommunikation	Informationen sach- und fachbezogen austauschen
Bewertung	Chemische Sachverhalte in verschiedenen Kontexten erkennen und bewerten

Das Anwenden experimenteller Methoden im Kompetenzbereich Erkenntnisgewinnung erfordert in der Regel Fachwissen, um unter Beachtung möglicher Reaktionsabläufe geeignete Experimente zu planen, Versuchsanleitungen auszuwählen und umzusetzen. Fachwissen ist auch notwendig, um einen chemischen Sachverhalt zu kommunizieren, unabhängig davon, ob dies in einem Text, einem Referat oder einem Gespräch innerhalb oder außerhalb des Unterrichts erfolgt. Gleichzeitig stellt das Fachwissen die Grundlage zum Bewerten chemischer Sachverhalte dar.

Um eine naturwissenschaftliche Grundbildung im Sinne von *scientific literacy* (Bybee 2002) zu erreichen, müssen alle Kompetenzbereiche im Schulunterricht gleichberechtigt berücksichtigt werden. Während die in diesem Zusammenhang relevanten Themen und Aspekte beispielsweise in den USA bereits vor Jahren im Rahmen des Projekts 2061 der American Association for the Advancement of Science (AAAS 1990; AAAS 1993) diskutiert wurden und schon frühzeitig vom National Research Council in die Standards (National Research Council 1996) aufgenommen wurden, sind viele dieser Überlegungen für das deutsche Bildungssystem teilweise neu. Der stark auf Inhaltsvermittlung ausgerichtete Unterricht muss entsprechend für bisher wenig berücksichtigte Kompetenzbereiche geöffnet werden. Eine Realisierung ist nur möglich, wenn gleichzeitig sowohl geeignete Lernaufgaben als auch Prüfungsaufgaben entwickelt werden, die nicht nur Fachwissen sondern auch die übrigen Kompetenzen abprüfen können. Trotz der oben beschriebenen engen Verknüpfung der Kompetenzbereiche müssen

die Aufgaben und Testinstrumentarien in der Lage sein, die Fähigkeiten der Schülerinnen und Schüler in den einzelnen Bereichen zu überprüfen, also Aussagen über den jeweils speziellen Kompetenzbereich zu machen, und müssen sich von anderen Kompetenzbereichen abgrenzen lassen (diskriminante Validität). Gerade hinsichtlich des Fachwissens scheint diese Abgrenzung nicht trivial.

2.1 Der Kompetenzbereich Erkenntnisgewinnung

Der Kompetenzbereich „Erkenntnisgewinnung“ umfasst neben wissenschaftstheoretischen Grundlagen, naturwissenschaftlicher Theorie- und Modellbildung auch naturwissenschaftliche (nicht nur) experimentelle Untersuchungen, also das Erwerben von Wissen über und das Umgehen mit experimentell-naturwissenschaftlichen Arbeitsweisen. Im deutschsprachigen Raum gibt es nur wenig Forschung, die sich mit der Vertrautheit von Lehrenden und Lernenden mit den Erkenntnismethoden der Naturwissenschaften befassen (Hößle/Höttecke/Kircher 2004; Höttecke 2001). Im angloamerikanischen Sprachraum wird diese Diskussion unter dem Stichwort *Nature of Science* geführt. Die amerikanischen Standards (AAAS 2001) unterscheiden hier zwischen *Scientific World View*, *Scientific Inquiry* und *Scientific Enterprise*. Das „*Scientific Inquiry Konzept*“ wird durch vier Facetten beschrieben:

- Evidence and Reasoning
- Scientific Investigations
- Scientific Theories
- Avoiding Bias in Science

Naturwissenschaftliche Untersuchungen und Arbeitsweisen gehören damit als *Scientific Investigations* in den Bereich *Scientific Inquiry*. In allen Fällen wird dieser Prozess auch als das problemorientierte Testen von Hypothesen mit experimentellen Methoden beschrieben (z.B. Bybee 2002; Chinn/Malhotra 2002; Duit/Gropengießer/Stäudel 2004a; Fischer 1998; Klahr 2000; Sandoval/Reiser 2004).

2.2 Naturwissenschaftlich-experimentelle Arbeitsweisen

Naturwissenschaftliche Arbeitsweisen können als Problemlösestrategien für ein naturwissenschaftliches Problem aufgefasst werden (Burns/Vollmeyer 2000; Klahr 2000). Klahr beschreibt in seinem Modell *Scientific Discovery as Dual Search* (Klahr 2000; Klahr/Dunbar 1988) den naturwissenschaftlichen Erkenntnisprozess mit Hilfe zweier Suchräume, dem Hypothesensuchraum (*Space of Hypothesis*) und dem Experimentiersuchraum (*Space of Experiments*). Im Hypothesensuchraum wird eine Hypothese gesucht und ausgewählt, im Experimentiersuchraum wird ein Experiment gesucht, das interpretierbare Ergebnisse liefert, um diese Hypothese zu testen. Der letzte wesentliche

Schritt ist nach Klahr auch zugleich der Anspruchsvollste. Die erhaltenen Ergebnisse müssen interpretiert und auf die Eingangshypothese bezogen werden, um gegebenenfalls über eine weitere Suche im Hypothesen- oder Experimentiersuchraum zu entscheiden bzw. auch vorhandene Überzeugungen zu revidieren. Die genannten Schritte lassen sich weiter aufgliedern, wobei an dieser Stelle auf Walpuski (2006) verwiesen wird, der weitere Differenzierungen zusammengefasst dargestellt.

Die hier thematisierten drei Schritte – Hypothesenbildung, Planung von Experimenten und Schlussfolgerung aus experimentellen Daten – werden nicht als der einzige Weg für naturwissenschaftliches Arbeiten verstanden (z.B. Bauer 1992; McComas 1996; Palincsar/Anderson/David 1993; Pfeifer 2003), aber als Hinweis auf einen möglichen Weg, der auf viele naturwissenschaftliche Probleme anwendbar ist. Dennoch wird gerade dieser Punkt von Lederman u.a. (2002) kritisch betrachtet, da bei Schülerinnen und Schülern der Eindruck entstehen könnte, dass es die „Allheilmethode“ zur Lösung naturwissenschaftlicher Probleme gibt. Aber auch Lederman schränkt die Methodik auf eine überschaubare Anzahl an Tätigkeiten ein, ohne jedoch eine wiederkehrende Abfolge von Tätigkeiten festzulegen. „It is true that scientists observe, compare, measure, test, speculate, hypothesize, create ideas and conceptual tools, and construct theories and explanations.“ (Lederman u.a. 2002, S. 501) In diesem Zusammenhang spricht Bauer (1992) von einer großen Spanne von Charakteristika, die naturwissenschaftliche Aktivitäten beschreiben. Trotzdem ist es für die Praxis in der Schule legitim, diese große Spanne einzugrenzen und für Lernende situationsadäquat auf diese Weise begreiflich zu machen. Man sollte dann aber auch Situationen nutzen, in denen man mehrere Ansätze zur Problemlösung fordert und im Unterricht nachvollzieht.

Auch in der deutschsprachigen Literatur für Lehrerinnen und Lehrer findet sich die Methode zum Lösen naturwissenschaftlicher Probleme entlang des beschriebenen Weges der Erkenntnisgewinnung wieder (z.B. Duit/Gropengießer/Stäudel 2004b; Pfeifer 2003; Schmidkunz/Lindemann 1992; Stäudel/Werber/Freimann 2002).

Erste Umsetzungen der aktuellen bildungspolitischen Entwicklung zeigen sich z.B. in dem Entwurf des Kernlehrplans Chemie (Lehrplangruppe 2007) in dem unter anderem verstärkt experimentell-naturwissenschaftliches Arbeiten thematisiert wird. Hier werden für den Kompetenzbereich Erkenntnisgewinnung beispielsweise nachstehende Kompetenzen festgelegt:

Schülerinnen und Schüler...

- beobachten und beschreiben Phänomene und Vorgänge und unterscheiden dabei Beobachtung und Erklärung
- erkennen und entwickeln Fragestellungen, die mit Hilfe chemischer und naturwissenschaftlicher Kenntnisse und Untersuchungen zu beantworten sind
- stellen Hypothesen auf, planen geeignete Untersuchungen und Experimente zur Überprüfung, führen sie unter Beachtung von Sicherheits- und Umweltaspekten durch und werten sie unter Rückbezug auf die Hypothesen aus
- interpretieren Daten, Trends, Strukturen und Beziehungen, erklären diese und ziehen geeignete Schlussfolgerungen

- recherchieren in unterschiedlichen Quellen (Print- und elektronische Medien) und werten die Daten, Untersuchungsmethoden und Informationen kritisch aus.

Während also in unterschiedlichen Kontexten verstärkt Kompetenzen im Bereich der naturwissenschaftlichen Erkenntnisgewinnung angestrebt und eingefordert werden, gibt es kaum Hinweise darauf, wie diese Kompetenzen bei Schülerinnen und Schülern nachgewiesen bzw. gemessen werden können. Im deutschsprachigen Raum gibt es neben den im folgenden dargestellten Arbeiten weitere erste Schritte zur Entwicklung von Paper-Pencil-Tests zur Messung naturwissenschaftlicher Arbeitsweisen mit dem Schwerpunkt auf biologisch-didaktischen Fragestellungen (Hammann 2004). Des Weiteren wird ein computerbasiertes Testinstrument mit dem Ziel, den fachspezifischen Anteil einer fachübergreifenden Kompetenz „Problemlösen“ für den Chemieunterricht zu erheben von Koppelt/Tiemann (2008) entwickelt. Außerdem ist ein Test zu Metakognitiven Strategien im naturwissenschaftlichen Bereich von Thillmann (2007) verfügbar, der Einschätzungen der Schülerinnen und Schüler abfragt.

3. Naturwissenschaftlicher-Arbeitsweisen-Test (NAW-Test)

Aufbauend auf der Theorie Klahrs (2000) entwickelten Kieren (2004) und Walpuski (2006) ein Testinstrument für die 7. Klasse zur Messung experimenteller Arbeitsweisen. Dabei wird davon ausgegangen, dass sich experimentell-naturwissenschaftliches Arbeiten in dieser Jahrgangsstufe im Wesentlichen auf die in Abbildung 1 aufgeführten zentralen Schritte Ideen-/ Hypothesenbildung, experimentelle Umsetzung und Schlussfolgerung vereinfachen lässt. Statt von Hypothesenbildung zu sprechen wird dieser Altersgruppe die Formulierung Ideenbildung gerechter, da die theoretische Fundierung dieser Annahmen wenig ausgeprägt ist.

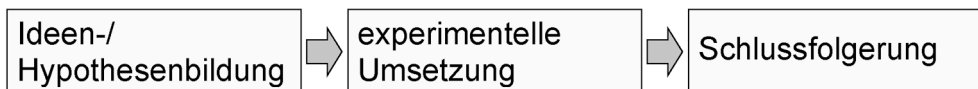


Abb. 1: Schritte des experimentell-naturwissenschaftlichen Arbeitens

Der Test ist im Multiple-Choice-Antwortformat konstruiert. Zunächst erhalten die Schülerinnen und Schüler einen Informationstext mit dem für die Aufgabe benötigten deklarativen Fachwissen um Vorwissenseinflüsse zu minimieren. Danach folgt eine Situations- oder Problembeschreibung. Es muss angegeben werden, ob die vorgegebenen Ideen/Hypothesen geeignet oder ungeeignet sind, um das Problem zu lösen, oder ob sie prüfbar oder unprüfbar sind. Eine Besonderheit des NAW-Tests besteht darin, dass die Testpersonen durch das dichotome Antwortformat bewusst einer Aussage zustimmen oder widersprechen müssen. Abbildung 2 zeigt zur Veranschaulichung ein Beispielitem.

1. Zuerst überlegen Manuel und Katrin, wodurch der komische Geschmack und Geruch des naturtrüben Apfelsafts verursacht wird.

Manuel und Katrin haben folgende Ideen. Welche Ideen lassen sich aus dem Text ableiten?

	ja	nein
Die Äpfel waren nicht richtig reif.	<input type="checkbox"/>	<input type="checkbox"/>
Der Apfelsaft hat begonnen zu gären.	<input type="checkbox"/>	<input type="checkbox"/>
Der Apfelsaft ist schlecht geworden.	<input type="checkbox"/>	<input type="checkbox"/>
Der Apfelsaft hat den Geschmack von anderen Lebensmitteln angenommen.	<input type="checkbox"/>	<input type="checkbox"/>
Der Apfelsaft enthält zu viel Fruchtzucker.	<input type="checkbox"/>	<input type="checkbox"/>

Abb. 2: Beispielitem NAW-Test

Im weiteren Verlauf des Tests muss der Proband auch die Angemessenheit vorgeschlagener Experimente einschätzen oder angeben, welche Schlussfolgerungen aus den erhaltenen Informationen gezogen oder verallgemeinert werden dürfen.

Die ursprüngliche Variante des NAW-Tests (Kieren 2004), die von Walpuski (2006) an einer größeren Stichprobe in verschiedenen Formen getestet wurde, erfüllt die klassischen Testgütekriterien (Objektivität, Reliabilität, Validität) nur eingeschränkt. Mit *internen Konsistenzen* von $.20 < \text{Cronbachs } \alpha < .50$ (je nach Testform) weist der Test eine zu hohe Heterogenität auf, als dass er verlässliche Aussagen zuließe. Zudem liegt die *mittlere Aufgabenschwierigkeit* zwischen 0.60 und 0.97 (je nach Testform), woraus zu schließen ist, dass der Test insgesamt zu leicht ist. Der Test wurde von Klos (2007) überarbeitet. Um die Testschwierigkeit zu erhöhen, wurden zunächst für jeden Aufgabenbereich weitere, plausible Distraktoren und zusätzliche Aufgaben zur *experimentellen Umsetzung* entwickelt. Während in der ursprünglichen Testform lediglich die *experimentelle Umsetzung* einer *geeigneten Idee/Hypothese* zu beurteilen war, gilt es nun auch die *experimentelle Umsetzung* einer zur Problemlösung *nicht geeigneten Idee/Hypothese* zu verfolgen. Dies bedeutet, dass der Proband eine Idee zwar als *ungeeignet* klassifizieren muss, deren *experimentelle Umsetzung* aber nichtsdestotrotz im weiteren Verlauf des Tests verfolgen und beurteilen soll. Hierdurch wurde eine Problematik aufgenommen, deren Wichtigkeit durch internationale Studien belegt ist. Bady (1979) und Lederman (1992) zeigen, dass es in Schulen versäumt wird, Falsifikationsexperimente durchzuführen. Stattdessen werden Bestätigungsexperimente gezeigt, anhand derer die Schülerinnen und Schüler nicht lernen können, Hypothesen zu falsifizieren.

3.1 NAW-Test Jahrgangsstufe 7

Der überarbeitete NAW-Test für die 7. Jahrgangsstufe bedarf einer Bearbeitungszeit von 20 Minuten und besteht aus 39 Items. Insgesamt fallen 15 Items in den Inhaltsbereich *Idee / Hypothese*, die sich auf drei Itemstämme verteilen. Für den Bereich *experimentelle Umsetzung* werden 14 Items in drei Itemstämme verwendet, während für den Bereich *Schlussfolgerung* zehn Items für zwei Itemstämme konstruiert wurden.

3.1.1 Testgüte des NAW-Tests (Jahrgangsstufe 7)

Um Angaben über die Aussagekraft des Tests machen zu können, wurden zunächst Reliabilitätsanalysen durchgeführt. Lienert/Raatz (1998) erachten eine Reliabilität zwischen 0.5 und 0.7 für ausreichend, wenn Gruppenvergleiche durchgeführt werden. Das Ergebnis der Reliabilitätsanalyse über den Gesamttest ist mit einem Cronbachs $\alpha = .73$ durchaus zufriedenstellend. Allerdings könnten die niedrigeren Reliabilitätswerte für die drei oben beschriebenen Teilskalen *Ideen-/Hypothesenbildung*, *experimentelle Umsetzung* und *Schlussfolgerung* darauf hinweisen, dass es sich um nicht isolierbare Einzelskalen handelt (vgl. Tab 2). Da in Cronbachs α jedoch die Skalenlänge eingeht, kann an dieser Stelle nur festgestellt werden, dass der Gesamttest homogener ist als die hypothetisch angenommenen Einzelskalen.

Tab. 2: Interne Konsistenz der Einzelskalen (α -Werte) im NAW-Test 7		
	jeweilige Itemzahl	errechnete α -Werte (N = 572)
Interne Konsistenz Gesamttest	39	.73
Skala Ideen-/Hypothesenbildung	15	.64
Skala Experimentelle Umsetzung	14	.54
Skala Schlussfolgerung	10	.49

Auch eine Faktorenanalyse bestätigt, dass mit dem NAW-Test *ein* Konstrukt erhoben wird. Bis zu 56 % der Varianz lassen sich hierbei durch einen Faktor aufklären, während die beiden anderen gefundenen Faktoren statistisch vernachlässigbar sind und *eine* Hauptkomponente extrahiert wird.

Gerade die Überprüfung der Validität ist in Bezug auf den NAW-Test außerordentlich bedeutsam, da davon ausgegangen wird, dass mit Hilfe des Testinstruments prozedurales Wissen im Bereich der experimentellen Arbeitsweisen getrennt vom Fachwissen erhoben werden kann. Da dieses Wissen anhand naturwissenschaftlicher Inhalte getestet wird oder darauf aufbaut, gilt es zu überprüfen, ob der NAW-Test sich im Sinne diskriminanter Validität empirisch vom Fachwissenstest trennen lässt. Zunächst kann davon ausgegangen werden, dass das Fachwissen, über welches die Schülerinnen und Schüler zu Beginn der 7. Jahrgangsstufe verfügen, eher gering ist, da der Chemieunter-

richt an Gymnasien erst mit Beginn dieser Klasse einsetzt. Aufgrund dessen ist es verhältnismäßig einfach ein Themengebiet zu finden, das auf der einen Seite an die Lebenswelt der Schülerinnen und Schüler anknüpft und es ihnen so ermöglicht, sich in die im Test beschriebene Problemstellung zu versetzen, auf der anderen Seite aber im bisherigen naturwissenschaftlichen Unterricht nicht konkret thematisiert wurde.

Zur Prüfung der Abhängigkeit des NAW-Tests vom Fachwissen wurden die NAW-Test-Ergebnisse der Datenerhebung mit den in der gleichen Studie erhobenen Daten des Fachwissenstests Chemie korreliert.

Zwischen dem NAW-Test und dem eingesetzten Fachwissenstest (Cronbachs $\alpha = .87$, 66 Items, Itemschwierigkeit $.17 < p < .67$) ergibt sich ein Korrelationskoeffizient von $r = .070$ ($p = .101$). Es handelt sich hierbei nach Bühl und Zöfel (2002) um eine sehr geringe Korrelation, die verdeutlicht, dass zwischen den mit den beiden Testinstrumenten erhobenen Konstrukten kein erwähnenswerter Zusammenhang besteht. Ebenso ist die Korrelation von $r = .360$ ($p < .001$) mit den im eingesetzten Kognitiven-Fähigkeiten-Test (KFT) (Heller/Perleth 2000) erreichten Ergebnissen gering (Bühl/Zöfel 2002), so dass die kognitiven Fähigkeiten zwar einen Einfluss auf die Leistung im NAW-Test haben, der NAW-Test sich aber durchaus empirisch von Intelligenzparametern trennen lässt. Auch der Einfluss der Lesefähigkeit, gemessen anhand der verbalen Subskalen des KF-Tests, ist mit einer geringen Korrelation von $r = .263$ ($p < .001$) vernachlässigbar. Somit kann davon ausgegangen werden, dass der NAW-Test weder hauptsächlich kognitive Fähigkeiten noch reines Fachwissen, sondern eine darüber hinaus vorhandene Komponente erfasst.

3.1.2 *Ausgewählte weitere Ergebnisse*

Der NAW-Test für die Jahrgangsstufe 7 wurde im Rahmen einer Untersuchung mit 600 Schülerinnen und Schülern eingesetzt (Klos, 2007). Hier zeigt sich ein überraschender Geschlechtereffekt. Ein Vergleich der Mädchen und Jungen hinsichtlich ihrer NAW-Testergebnisse führt zu deutlichen Unterschieden zugunsten der Mädchen. Sie liegen mit einem Mittelwert von 23.28 Punkten ($SD = 5.232$) signifikant ($p < .001$) über dem Mittelwert der Jungen von 21.32 Punkten ($SD = 5.434$). Die Effektstärke liegt hier bei $\eta^2 = .152$ ($F_{(1; 569)} = 13.124$; $\eta^2 = .023$). Demgegenüber zeigt ein Vergleich der Mädchen (103.06 Punkte, $SD 32.10$) und Jungen (110.92 Punkte, $SD 35.93$) im Fachtest Chemie mit einer Signifikanz von $p = .011$ ($F_{(1; 547)} = 6.566$; $\eta^2 = .012$) den bekannten Vorteil zu Gunsten der Jungen. Diese unterschiedlichen Geschlechtereffekte deuten ebenfalls darauf hin, dass der NAW-Test eine andere Kompetenz misst als ein Fachwissenstest.

3.2 *NAW-Test Oberstufe*

Im Vergleich zum NAW-Test für die 7. Jahrgangsstufe, in dem das Vorwissen der Probanden als gering eingestuft werden kann und es entsprechend einfach ist, das zum Be-

antworten der Testitems benötigte Fachwissen bereitzustellen, ist dieser Aspekt für einen Oberstufenkurs ungleich schwerer umsetzbar (Henke 2007). Im Gegensatz dazu kann davon ausgegangen werden, dass Schülerinnen und Schüler nach mindestens vier Jahren Chemieunterricht bis zur Oberstufe unterschiedliche Vorkenntnisse haben und damit über das für die Beantwortung eines NAW-Test-Items notwendige Fachwissen bereits verfügen oder aber sich dieses neu erarbeiten müssen. Da Vorwissen einen wesentlichen Prädiktor für Schülerleistungen darstellt (Ericsson/Crutcher 1990; Weinert/Helmke/Schneider 1990), folgt daraus, dass gerade solchen Schülerinnen und Schülern, die ein themenspezifisches hohes Vorwissen aufweisen, die Beantwortung der NAW-Test-Items leichter fallen wird und dieser Umstand bei der Entwicklung des NAW-Test berücksichtigt werden muss.

Während der NAW-Test für die Jahrgangsstufe 7 als Multiple-Choice-Test konstruiert ist, enthält der NAW-Test für die Oberstufe neben solchen Items mit dichotomen Antwortformat auch einige Aufgaben mit offenem Antwortformat und bedarf einer Bearbeitungszeit von 60 Minuten. Des Weiteren werden hier nicht mehr wie beim NAW 7 alle drei wesentlichen Bereiche des experimentell-naturwissenschaftlichen Arbeitens in *einem*, sondern in *verschiedenen* Kontexten abgefragt. Insgesamt verteilen sich 24 Items auf die drei bekannten Inhaltsbereiche *Hypothesenbildung*, *experimentelle Umsetzung* und *Schlussfolgerung*.

3.2.1 Testgüte des NAW-Tests (Jahrgangsstufe 12)

Die Überprüfung der internen Konsistenz mit Hilfe von Reliabilitätsanalysen bestätigt die Ergebnisse der entsprechenden Analysen für den NAW 7. Auch im NAW 12 liegen die α -Werte für den Gesamttest deutlich höher als die α -Werte der drei vermuteten Subskalen (vgl. Tab. 3). Die Daten wurden hier aus zwei verschiedenen Messzeitpunkten gewonnen.

Tab. 3: Interne Konsistenz der Einzelskalen (α -Werte) im NAW-Test 12			
	jeweilige Itemzahl	1. Messzeitpunkt N = 139	2. Messzeitpunkt N = 139
Interne Konsistenz Gesamttest	24	.82	.86
Skala Ideen-/Hypothesenbildung	9	.71	.74
Skala Experimentelle Umsetzung	5	.69	.72
Skala Schlussfolgerung	10	.67	.70

Eine explorative Faktorenanalyse zeigt ebenfalls keine Hinweise für die drei vermuteten Skalen. Während sich die *Skala Experiment* noch am besten abbildet und von den beiden anderen unterscheiden lässt, lassen sich die beiden anderen vermuteten Subskalen *Hypothese* und *Schlussfolgerung/Datenauswertung* nicht mehr trennscharf voneinander

unterscheiden. Die in diesem Zusammenhang von Hammann (2004) angenommenen Unterschiede der Fähigkeiten in den verschiedenen Skalen im Sinne unterschiedlicher Kompetenzniveaus können von Henke (2007) nicht bestätigt werden. Die Ergebnisse deuten vielmehr darauf hin, dass die theoretisch angenommenen drei Teilbereiche experimentell-naturwissenschaftlichen Arbeitens sich empirisch nicht einzeln abbilden lassen und daher als *eine* Fähigkeit angenommen werden können.

Da für den NAW-Test 12 ein im Vergleich zur 7. Klasse stärkerer Einfluss der Vorwissens nicht a priori ausgeschlossen werden kann, ist hier ebenfalls zu überprüfen, ob sich die NAW-Testergebnisse von denen anderer Testinstrumente zur Erhebung fachlicher Inhalte unterscheiden lassen und somit Fähigkeiten anderer Kompetenzbereiche erfasst werden. Tatsächlich zeigt sich hier eine Korrelation von $r = .68$ (mittlere Korrelation; $p < .000$) zwischen den angenommenen Konstrukten NAW und Fachwissen. Um diesem theoretisch nicht angenommenen Zusammenhang der empirischen Daten weiter nachzugehen, wurde eine konfirmatorische Faktorenanalyse (CFA) durchgeführt (Bollen 1989), die zeigt, inwieweit der eingesetzte Fachwissenstest und der NAW-Test 12 dasselbe hypothetisch angenommene Konstrukt messen. In einem ersten Schritt wurde ein Modell aufgestellt, in dem die beiden Konstrukte *Fachwissen* und *experimentell-naturwissenschaftliche Arbeitsweisen* als zwei getrennte latente Variablen definiert werden. Eine Modellprüfung zeigt, dass eine hypothetische Gleichsetzung der Variablen *Fachwissen* und *experimentell-naturwissenschaftliche Arbeitsweisen* in dem „genesteten Modell“ (Abb. 4) zu einer schlechteren Passung führt als das anhand der empirischen Daten errechnete, freie „a-priori-Modell“ (Abb. 3).

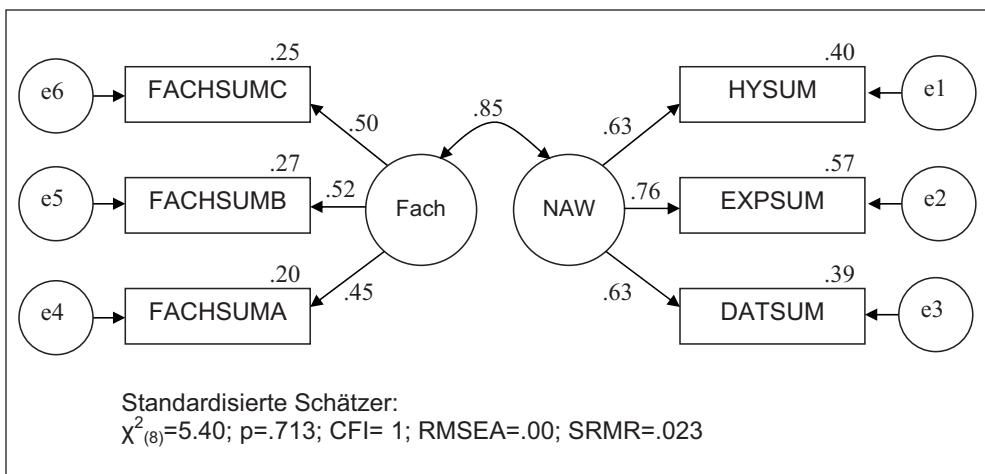


Abb. 3: Angenommenes „a-priori-Modell“ mit Angabe der Fit-Indices

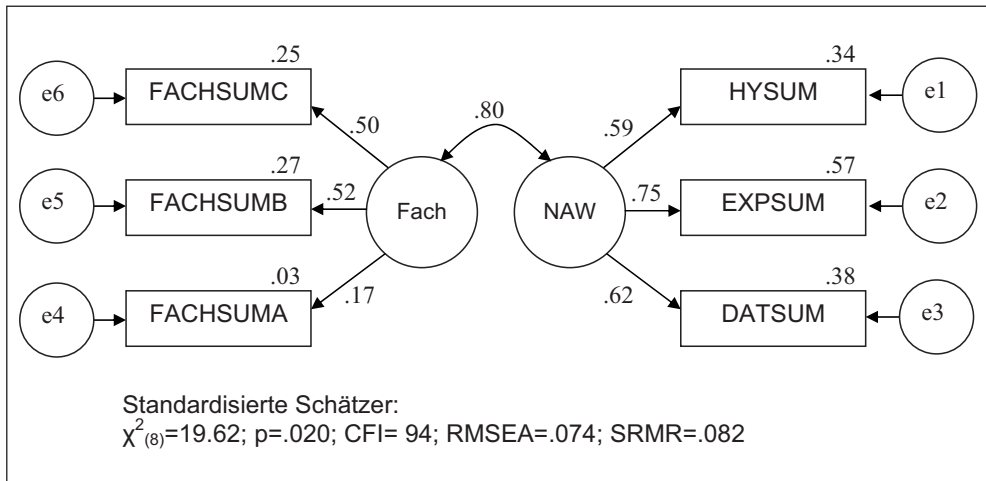


Abb. 4: Angenommenes „genestetes“ Modell mit Angabe der Fit-Indices

Der Differenzwert der Chi-Quadrat-Werte zwischen beiden Modellen beträgt $\Delta\chi^2_{(1)} = 14.215$ ($p < .001$) und der Differenzwert des CFI ist größer als .01. Beide Differenzwerte bestätigen die deutlich schlechtere Passung des *genesteten* Modells auf die Datenstruktur (Cheung/Rensvold 2002). Die Fit-Indices, die sich deutlich verändert haben, zeigen ebenfalls, dass es sich nicht um zwei identische Variablen handelt. Insgesamt kann demnach davon ausgegangen werden, dass die beiden Testinstrumente *Fachtest* und *NAW-Test 12* unterschiedliche Konstrukte messen. Es handelt sich beim *NAW-Test 12* zwar um ein Testinstrument, zu dessen Bearbeitung Fachwissen benötigt wird, das aber darüber hinaus weitere Fähigkeiten misst. Da darüber hinaus gezeigt werden kann, dass zwischen Intelligenzparametern (KFT) nur geringe bis mittlere Korrelationen ($r = .45$, $p < .000$) bestehen, wird davon ausgegangen, dass mit dem *NAW-Test 12* ebenfalls prozedurales Wissen hinsichtlich experimentell-naturwissenschaftlichen Arbeiten erfasst wird, die nur teilweise mit Fachwissen und kognitiven Fähigkeiten verknüpft sind.

3.2.2 Ergebnisse der Videovalidierung

Das Problem der Validierung des *NAW-Tests* liegt im Fehlen eines Tests, der auch nur ansatzweise das prozedurale Wissen von Schülerinnen und Schülern über naturwissenschaftliche Arbeitsweisen erfasst. Aus diesem Grund wurde die Konstruktvalidierung mittels einer Videofallstudie mit 12 Schülerinnen und Schülern durchgeführt, in der die Lösungsprozesse einzelner Lernender im konkreten Problemlöseprozess aufgezeichnet und ausgewertet werden. Die so erhaltenen Verhaltensdaten der Schülerinnen und Schüler werden im Anschluss mit den Ergebnissen im *NAW-Test* korreliert.

Alle an der Videostudie teilnehmenden Schülerinnen und Schüler erhielten die gleiche Aufgabe in einer Interaktionsbox (Rumann, 2005; Walpuski 2006). Interaktionsboxen enthalten eine konkrete Problemstellung und Materialien, die für die Bearbeitung des Problems benötigt werden und mehrere Lösungswege ermöglichen, sowie Infokarten zu bestimmten benötigten fachlichen Inhalten. Auf diese Weise werden rezeptartige Versuchsvorschriften vermieden. Die Schülerinnen und Schüler müssen selbst das zur Lösung des Problems passende Experiment entwerfen und durchführen. Die Interaktionsbox dient hier als Diagnoseinstrument für die Fähigkeit der Lernenden, in einer Problemlösesituation experimentell-naturwissenschaftlich zu arbeiten (Henke 2007).

Die Schülerinnen und Schüler hatten die Aufgabe, die in der Interaktionsbox gegebene Problemstellung innerhalb von 30 Minuten zu bearbeiten. Während der Bearbeitung wurden die Aktivitäten mit einer Videokamera aufgezeichnet. Vor der Bearbeitung des Problems wurden die Schülerinnen und Schüler aufgefordert, laut zu denken (Spörring 1989), also möglichst alle Gedanken laut zu äußern. In dem Raum, in dem die Videoaufzeichnung stattfand, war außer dem Schüler oder der Schülerin nur noch eine Person zugegen, die für die Bedienung der Videokamera zuständig war und die Aufgabe hatte, die jeweilige Testperson an das „laute Denken“ zu erinnern, wenn ersichtlich war, dass eine Aktion ohne vorherige sprachliche Äußerung erfolgte.

Die Auswertung der Videos erfolgt mit Hilfe von Kodierschemata zur Erstellung von Lernprozessgrafiken, die zur Auswertung von experimentellen Arbeitsweisen in Videos aus Klassen- und Kleingruppensituationen eingesetzt wurden (Dettloff 2003; Rumann 2005; Walpuski 2006). Der Problemlöseverlauf wird hierbei zunächst grafisch visualisiert. Die Beschreibung und Berechnung der Ergebnisse beruht im Anschluss auf diesen Grafiken. Sie dienen hier der Erhebung von Verhaltensdaten einzelner Lernender in einer konkreten Problemlösesituation.

In den Lernprozessgrafiken sind die einzelnen Lösungsschritte (Hypothesen, Experimente, Schlussfolgerungen) mit Hilfe verschiedener Symbole unterscheidbar. Durch Verbindungslinien zwischen den einzelnen Symbolen werden Lösungsverläufe dargestellt. So kann entschieden werden, ob verschiedene Lösungsansätze untereinander verbunden sind, wenn sich z.B. eine Hypothese auf ein Experiment eines anderen Lösungsansatzes bezieht oder Schlussfolgerungen aus verschiedenen Lösungsansätzen zu einer neuen Schlussfolgerung vereinigt werden. Aufgrund der Verbindungslinien kann auch entschieden werden, ob die einzelnen Schritte auf planvollem, theoriegeleitetem Handeln beruhen und ob aus den einzelnen Lösungsschritten Konsequenzen für das weitere Arbeiten gezogen werden. Durch die Kennzeichnung falscher Lösungsschritte ist eine detaillierte Fehleranalyse möglich. Hierzu zählt beispielsweise die Beantwortung nachstehender Fragen: Ist der Schüler bzw. die Schülerin zu einer Lösung des Problems gekommen? Hat er auf seinem Lösungsweg Fehler begangen? Ist er oder sie im Sinne der naturwissenschaftlichen Erkenntnisgewinnung logisch vorgegangen?

Die aus den zwölf Videos erhaltenen Daten werden mit den Ergebnissen des NAW-Tests korreliert, der die Fähigkeit der Schülerinnen und Schüler messen soll, ein naturwissenschaftliches Problem experimentell zu bearbeiten. Deshalb wird als Kriterium die auf die jeweilige Gesamtzahl an durchgeführten Problemlösesequenzen bezogene Zahl

richtiger Sequenzen gewählt. Da sowohl die Werte im NAW-Test als auch die Anzahl der relativen richtigen Problemlösesequenzen aufgrund der kleinen Probandenzahl ($N=12$) nicht als bivariat normalverteilt angenommen werden können, wird für die Ermittlung des vermuteten Zusammenhangs zwischen beiden Werten die non-parametrische Rangkorrelation mit dem Korrelationskoeffizienten Spearman- ρ gewählt (Bortz/Lienert/Boehnke 2000). Hier findet man eine mittlere Korrelation ($\rho = .645$, $p = .012$) zwischen den Ergebnissen im NAW-Test und den richtigen Problemlösesequenzen, die aus den Videos bestimmt wurden.

Dieser Zusammenhang zwischen der Verhaltensweise der Testperson in einer realen Problemlösesituation und den im NAW-Test erreichten Ergebnisse kann als weiteres Indiz für die Validität des Testinstruments angesehen werden.

3.2.3 Ausgewählte weitere Ergebnisse

Der NAW-Test für die Jahrgangsstufe 12 wurde in einer anderen Studie (Henke 2007) parallel zu einem Fachwissenstest zu zwei verschiedenen Messzeitpunkten (Pre- und Post-Erhebung) im Abstand eines Jahres in derselben Stichprobe aber in zwei verschiedenen, aufgrund der Kontrolldaten vergleichbaren Lerngruppen (Kontrollgruppe und Experimentalgruppe) eingesetzt. Zum Vergleich der Zuwächse in beiden Gruppen wurde eine univariate ANOVA mit Messwiederholung gerechnet, in die das Fachwissen als Kovariate einging. Beide Gruppen zeigen einen höheren Testmittelwert im Post-Test (vgl. Tab. 4).

Tab. 4: Mittelwerte im NAW-Test				
	Mittelwert (Pre)	SD	Mittelwert (Post)	SD
Gruppe 1	53.9	10.7	71.8	7.0
Gruppe 2	52.5	13.1	58.2	11.7

Dies spiegelt sich auch im berechneten Haupteffekt ($F_{(1;36)} = 52,12$; $p < .001$; $\eta^2 = .59$) für den Messzeitpunkt wider. Das bedeutet, dass die Schülerinnen und Schüler beider Gruppen in dem Jahr zwischen den beiden Messzeitpunkten einen Lernzuwachs in ihren experimentell-naturwissenschaftlichen Fähigkeiten erlangt haben. Die Experimentalgruppe zeigt im Mittel einen höheren Lernzuwachs als die Kontrollgruppe. Dieser Interaktionseffekt ($F_{(1;36)} = 13,93$; $p < .01$; $\eta^2 = .28$) zeigt sich ebenfalls in der durchgeführten Varianzanalyse. Der NAW-Test ist also in der Lage, gruppenspezifische Zuwächse der experimentell-naturwissenschaftlichen Fähigkeiten zwischen zwei Messzeitpunkten zu bestimmen.

4. Zusammenfassung

Durch PISA wurde die Diskussion über das deutsche Bildungssystem neu angestoßen. Durch die Vergleichsstudien zeigte sich, dass gerade Fähigkeiten wie experimentell-naturwissenschaftliches Arbeiten höhere Aufmerksamkeit und Förderung gelten muss. In diesem Beitrag wird die theoriegeleitete Entwicklung eines Testinstruments entwickelt, das experimentell-naturwissenschaftliche Fähigkeiten von Schülerinnen und Schülern erfasst. Es konnte gezeigt werden, dass das entwickelte Testinstrument, der Naturwissenschaftliche Arbeitsweisentest, eine Kompetenz misst, die sich sowohl vom reinen chemischen Fachwissen als auch von den kognitiven Fähigkeiten der Schülerinnen und Schüler unterscheidet. Varianten dieses Tests sind für die Jahrgangsstufen 7 und 12 an Gymnasien konstruiert, evaluiert und eingesetzt worden. Umgekehrt ermöglicht diese erfolgreiche Testentwicklung auch die Aussage, dass es sich bei den bisher überwiegend normativ beschriebenen Kompetenzen Fachwissen und experimentelle Erkenntnisgewinnung tatsächlich um verschiedene Kompetenzen handelt.

Literatur

- AAAS (1990): *Science for All Americans*. New York: Oxford University Press.
- AAAS (1993): *Benchmarks for Science Literacy*. New York: Oxford University Press.
- AAAS (2001): *Atlas of Science Literacy*. New York/Oxford: Oxford University Press.
- Bady, R.J. (1979): Students' Understanding of the Logic of Hypothesis Testing. In: *Journal of Research in Science Teaching* 16, pp. 61–65.
- Bauer, H.H. (1992): *Scientific Literacy and the Myth of Scientific Method*. Urbana/Chicago: University of Illinois Press.
- Baumert, J./Bos, W./Lehmann, R. (2000b): TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Band 1: Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit. Opladen: Leske + Budrich.
- Baumert, J./Bos, W./Lehmann, R. (2000a): TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Band 2: Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe. Opladen: Leske + Budrich.
- Baumert, J./Klieme, E./Neubrand, M./Prenzel, M./Schiefele, U./Schneider, W./Stanat, P./Tillmann, K.-J./Weiß, M.H. (2001): *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.
- Baumert, J./Lehmann, R./Lehrke, M./Schmitz, B./Clausen, M./Hosenfeld, I./Köller, O. (1997): *TIMSS. Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde*. Opladen: Leske + Budrich.
- Bollen, K.A. (1989): *Structural equations with latent variables*. New York: Wiley & Sons.
- Bortz, J./Lienert, G./Boehne, K. (2000): *Verteilungsfreie Methoden in der Biostatistik*. Berlin: Springer-Verlag.
- Bühl, A./Zöfel, P. (2002): *SPSS 11. Einführung in die moderne Datenanalyse unter Windows*. München: Pearson Studium.

- Burns, B.D./Vollmeyer, R. (2000): Problem Solving: Phenomena in Search of a Thesis. In Gleitmann, L./Joshi, A. K. (Hrsg.): Proceedings of the Twenty-second Annual Meeting of the Cognitive Science Society. Hillsdale: Lawrence Erlbaum, pp. 627–632.
- Bybee, R.W. (2002): Scientific Literacy – Mythos oder Realität? In Gräber, W./Nentwig, P./Koballa, T./Evans, R. (Hrsg.): Scientific Literacy. Opladen: Leske + Budrich, S. 21–43.
- Cheung, G.W./Rensvold, R.B. (2002): Evaluating goodness-of-Fit Indexes for Testing Measurement Invariance. In: Structural Equation Modeling 9(2), pp. 233–255.
- Chinn, C.A./Malhotra, B.A. (2002): Epistemologically Authentic Inquiry in School: A Theoretical Framework for Evaluating Inquiry Tasks. In: Science Education 86, pp. 175–218.
- Dettloff, C. (2003): Videoanalyse von Frontal- und Kleingruppenunterricht am Beispiel einer Problemstellung zum Themenbereich Säure-Base. Universität Essen 1. Staatsexamensarbeit.
- Deutscher Verein zur Förderung des mathematischen und naturwissenschaftlichen Unterrichts (2003): Lernen und Können im naturwissenschaftlichen Unterricht. Denkanstöße und Empfehlungen zur Entwicklung von Bildungs-Standards in den naturwissenschaftlichen Fächern Biologie, Chemie und Physik (Sekundarstufe I).
- Duit, R./Gropengießer, H./Stäudel, L. (2004a): Naturwissenschaftliches Arbeiten – eine Einführung. In Duit, R./Gropengießer, H./Stäudel, L. (Hrsg.): Naturwissenschaftliches Arbeiten: Unterricht und Material 5-10. Seelze-Velber: Friedrich-Verlag, S. 4–8.
- Duit, R./Gropengießer, H./Stäudel, L. (2004b): Naturwissenschaftliches Arbeiten: Unterricht und Material 5-10. Seelze-Velber: Friedrich-Verlag.
- Ericsson, K.A./Crutcher, R.J. (1990): The Nature of Exceptional Performance. In Baltes, P. B./Featherman, D.L./ Lerner, R.M. (Hrsg.): Life-span development and behavior. Hillsdale, NJ: Academic Press, pp. 187–217.
- Fischer, H.E. (1998): Scientific Literacy und Physiklernen. In: Zeitschrift für Didaktik der Naturwissenschaften 4, S. 41–52.
- Hammann, M. (2004): Kompetenzentwicklungsmodelle. Merkmale und ihre Bedeutung – dargestellt anhand von Kompetenzen beim Experimentieren. In: Der mathematische und naturwissenschaftliche Unterricht 57, S. 196–203.
- Heller, K.A./Perleth, Ch. (2000): Kognitiver Fähigkeitstest für 4. bis 12. Klassen Revision (KFT 4-12+R). Göttingen: Hogrefe.
- Helmke, A. (2000): TIMSS und die Folgen: Der weite Weg von der externen Evaluation der Schulleistungen zur Verbesserung des Lehrens und Lernens. In: Trier, U.P. (Hrsg.): Bildungswirksamkeit zwischen Forschung und Politik. Chur: Ruegger, S. 135–164.
- Henke, C. (2007): Experimentell-naturwissenschaftliche Arbeitsweisen in der Oberstufe. Untersuchung am Beispiel des HIGHSEA-Projekts in Bremerhaven. Studien zum Physik- und Chemielernen. Berlin: Logos Verlag.
- Höfle, C./Höttecke, D./Kircher, E.H. (2004): Lehren und lernen über die Natur der Naturwissenschaften. Baltmannsweiler: Schneider Verlag Hohengehren.
- Höttecke, D. (2001): Die Natur der Naturwissenschaften historisch verstehen. Fachdidaktische und wissenschaftshistorische Untersuchungen. Studien zum Physiklernen. Berlin: Logos Verlag.
- Kieren, C. (2004): Naturwissenschaftliches Arbeiten im Anfangsunterricht – Entwicklung eines Testverfahrens. Schriftliche Hausarbeit im Rahmen der ersten Staatsprüfung für das Lehramt für die Sekundarstufe I/II. Universität Duisburg-Essen.
- Klahr, D. (2000): Exploring Science. The Cognition and Development of Discovery Processes. Massachusetts: Institute of technology.
- Klahr, D./Dunbar, K. (1988): Dual Space Search During Scientific Reasoning. In: Cognitive Science 12, S. 1–48.
- Klieme, E./Avenarius, H./Blum, W./Döbrich, P./Gruber, H./Prenzel, M./Reiss, K./Riquarts, K./Rost, J./Tenorth, H.-E./Vollmer, H. J. (2003): Zur Entwicklung nationaler Bildungsstandards – Eine Expertise. Bonn: Bundesministerium für Bildung und Forschung.

- Klos, S. (2007): Kompetenzförderung im naturwissenschaftlichen Anfangsunterricht – Der Einfluss eines integrierten Unterrichtskonzepts. Eingereichte Dissertation. Universität Duisburg-Essen.
- Koppelt, J./Tiemann, R.: Modellierung dynamischer Problemlösekompetenz im Chemieunterricht. Naturwissenschaftlicher Unterricht im internationalen Vergleich (im Druck).
- Koppelt, J./Tiemann, R. (2008): Modellierung dynamischer Problemlösekompetenz im Chemieunterricht. In: Höttecke, D. (Hrsg.): Kompetenzen, Kompetenzmodelle, Kompetenzentwicklung. Berlin: Lit Verlag, S. 362–363.
- Lederman, N.G. (1992): Students' and Teachers' Conceptions of the Nature of Science: A review of the Research. In: Journal of Research in Science Teaching 29, pp. 331–359.
- Lederman, N.G./Abd-El-Khalick, F./Bell, R.L./Schwartz, R.S. (2002): Views of Nature of Science Questionnaire: Toward Valid and Meaningful Assessment of Learners' Conceptions of Nature of Science. In: Journal of Research in Science Teaching 39, pp. 497–521.
- Lehrplangruppe (2007). Kernlehrpläne. URL: <http://db.learnline.de/angebote/kernlehrplaene/einfuehrung.jsp> [On-line].
- Lienert, G./Raatz, U. (1998): Testaufbau und Testanalyse. Weinheim: Psychologie Verlags Union.
- McComas, W. (1996): Ten myths of science: Reexamining what we think we know about the nature of science. In: School Science Mathematics 96, S. 10–16.
- Melle, I./Parchmann, I./Sumfleth, E. (2004): Kerncurriculum Chemie – Ziele, Rahmenbedingungen und Ansatzpunkte. In: Tenorth, H.-E. (Hrsg.): Kerncurriculum Oberstufe II. Biologie, Chemie, Geschichte, Politik. Weinheim: Beltz Verlag, S. 85–147.
- National Research Council (1996): National Science Education Standards. Washington, D.C.: National Academic Press.
- Palincsar, A.S./Anderson, C./David, Y.M. (1993): Pursuing Scientific Literacy in the Middle Grades through Collaborative Problem Solving. In: The Elementary School Journal 95, pp. 643–658.
- Pfeifer, P. (2003): Was heißt „naturwissenschaftliches Arbeiten“? In: Naturwissenschaften im Unterricht Chemie 14, S. 7–11.
- Pfeifer, P. (2003): Was heißt „naturwissenschaftliches Arbeiten“? In: Naturwissenschaften im Unterricht Chemie 14, 76/77, S. 7–11.
- Prenzel, M./Baumert, J./Blum, W./Lehmann, R./Leutner, D./Neubrand, M./Pekrun, R./Rolff, H.-G./Rost, J./Schiefele, U. (2004): PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs. Münster: Waxmann Verlag.
- Rumann, S. (2005): Kooperatives Arbeiten im Chemieunterricht. Entwicklung und Evaluation einer Interventionsstudie zur Säure-Base-Thematik. Studien zum Physik- und Chemielernen. Berlin: Logos Verlag.
- Sandoval, W.A./Reiser, B.J. (2004): Explanation-Driven Inquiry: Integrating Conceptual and Epistemic Scaffolds for Scientific Inquiry. In: Science Education 88, pp. 345–372.
- Schmidkunz, H./Lindemann, H. (1992): Das forschend-entwickelnde Unterrichtsverfahren – Problemlösen im naturwissenschaftlichen Unterricht. Magdeburg: Westarp Wissenschaften.
- Spörling, W. (1989): Qualitative Sozialforschung. Stuttgart: Teubner.
- Stäudel, L./Werber, B./Freimann, T. (2002): Naturwissenschaften – Verstehen & Anwenden. Seelze-Velber: Friedrich Verlag.
- Thillmann, H. (2007): Selbstreguliertes Lernen durch Experimentieren: Von der Erfassung zur Förderung. Dissertation an der Universität Duisburg-Essen, Fachbereich Psychologie.
- Walpuski, M. (2006): Optimierung von experimenteller Kleingruppenarbeit durch Strukturierungshilfen und Feedback. Eine empirische Studie. Studien zum Physik- und Chemielernen. Berlin: Logos Verlag.
- Weinert, F.E. (2001): Vergleichende Leistungsmessung in Schulen – Eine umstrittene Selbstverständlichkeit. In: Weinert, F. E. (Hrsg.): Leistungsmessungen in Schulen. Weinheim: Beltz, S. 17–31.

Weinert, F.E./Helmke, A./Schneider, W. (1990): Individual differences in learning performance and school achievement: Some plausible parallels and unexplained discrepancies. In: Learning and Instruction 2, S. 461–479.

Abstract: *The authors report on the development and evaluation of a method of testing aimed at measuring working methods in scientific experiments, the so-called NAW-Test. Two variants were developed for grades 7 and 12 and then employed in different studies, selected results of which are presented here. These results corroborate that the NAW-Test constitutes a valid testing method measuring a competence which is to be clearly distinguished from specialized subject-related knowledge.*

Anschrift der Autoren:

Silke Klos/Corinna Kieren/Dr. rer. nat. Maik Walpuski/Prof. Dr. rer. nat. Elke Sumfleth
Universität Duisburg-Essen, Campus Essen, Didaktik der Chemie, Schützenbahn 70, 45127 Essen
E-Mail: silke.klos@uni-due.de/corinna.kieren@uni-due.de/maik.walpuski@uni-due.de/elke.sumfleth@uni-due.de

Dr. rer. nat. Christian Henke, Studienseminar für Lehrämter an Schulen – Duisburg,
Bismarckstr. 120, 47057 Duisburg 70, E-Mail: christian.henke@uni-due.de